

Probabilistic Primary Vertex Selection

Ariel Schwartzman
Meenakshi Narain

November 7, 2002

Abstract

This notes addresses the issue of problems with primary vertex selection in RunIIa data reconstructed with DØ reco versions up to p11.13. We currently find many “*split vertices*”, mostly due to two tracks with poor resolution coming from the hard interaction and propose a new algorithm to solve this problem. In addition, a new probabilistic algorithm for Primary Vertex selection is presented.

Contents

1	Introduction	2
2	Primary Vertices in p11 Data	2
3	Minimum Bias Vertex Probability	3
4	Performance in the Monte Carlo	5
5	Performance in the Data	6
6	Conclusions	7

1 Introduction

Within the infrastructure of the DØ primary vertex reconstruction software, the last step is the identification of the hard scatter and additional minimum bias vertices of the event. The identified primary vertex is used for DØ reco to reconstruct jets, b-jets, electrons and missing transverse energy so it is important to optimize its performance to be able to efficiently reconstruct physics objects in the detector.

The vertex selection in the p11 version of DØ reco selects the vertex with the highest $\sum \log(p_T)$ [1] where the sum is over all tracks attached to the vertex.

The analysis of the first RunII data shows that poorly reconstructed tracks usually don't get attached to the primary vertex and tend to form "*split vertices*": low multiplicity vertices due to the hard interaction very near the higher multiplicity primary vertex. Given the fact that $\log(p_T)$ weighs the high p_T tracks more compared to the lower p_T ones, if one of the highest p_T tracks of the interaction gets attached to any "*split vertex*", the vertex selection algorithm might select the *splitv* as the primary vertex of the event.

In this note we propose a modification to the vertex selection algorithm to remove splitted vertices and discriminate primary from minimum bias vertices based on the probability that the vertex is consistent with a minimum bias interaction.

2 Primary Vertices in p11 Data

Figures 1 and 2 show the distribution of the number of attached tracks to all (selected) vertices for run 157708 where we note that there is a significant number of primary vertices made of 2 or 3 tracks. Figure 3 shows the distance in z from the selected primary vertex and all other vertices in the event indicating that there is a large fraction of vertices coming from the same interaction.

In order to gain better understanding of this feature, that instead of finding an unique primary vertex, the vertex reconstruction finds several "*split vertices*", we investigated properties of all the vertices found within 2 cm of the highest multiplicity interaction vertex.

For this study we first clustered tracks in the longitudinal plane and chose the track-cluster with the highest $\sum \log(p_T)$ as the primary interaction. We then define the primary vertex of the event as the vertex with the highest multiplicity within 2cm of the z of the selected

cluster of tracks. All remaining vertices from the same interaction were considered “*split vertices*”.

Figure 4 shows the χ^2 distribution for primary and “*split vertices*”. We note that the “*split vertices*” have a very broad χ^2 distribution showing that they are made of poorly reconstructed tracks. Figure 5 shows the number of attached tracks in primary and splitted vertices. It is clear that the majority of the “*split vertices*” have a multiplicity of two and are the ones populating the first bin of the figures 1 and 2.

Since *split* and primary vertices come from the same interaction, jet and missing energy reconstruction will not be significantly affected if one of the “*split vertex*” is selected as primary vertex. It may cause degradation of the missing energy resolution. The selection of a “*split vertex*” becomes a problem for secondary vertex reconstruction and identification of jets originating from b-quarks due to degradation of the track impact parameter and secondary vertex decay resolution.

Split vertices can be selected as primary vertices when the highest p_T track of the event get attached to them. When this happens, the $\sum \log(p_T)$ of the “*split vertex*” tracks might become larger than the same quantity for the good vertex.

In order to avoid the selection of “*split vertices*” we propose the following selection algorithm:

1. z-vertex-clustering: Cluster reconstructed vertices in z (Select cluster of vertices within 2 cm of each other)
2. For every cluster, select the highest multiplicity vertex and store it in the list of “selected” vertices. (this step removes “*split vertices*” from the list of selected vertices)
3. For every selected vertex, use all tracks within some distance around the selected vertex to compute the probability that the vertex comes from a Minimum Bias (MB) interaction.
4. Select the vertex with the smallest MB probability as the primary vertex.

The next section describes the definition of the MB vertex probability.

3 Minimum Bias Vertex Probability

Our goal here is to assign to each vertex, the probability that they come from a Minimum bias interaction.

We could use a multivariate technique to assign a probability to each vertex such as training a neural network with kinematic variables associated to hard and MB vertices. This, however, requires a model for the hard process which is physics dependent leading to an approach that will be very different for different physics process, and necessitate defining different probabilities for different physics process.

Instead, we propose to base our definition only on the properties of MB vertices and quantify how similar are the selected vertices to the MB kinematics. The only assumption is that tracks from hard interactions have higher p_T than tracks from MB interactions. Figure 6 shows the track p_T spectrum of minimum bias and primary interactions tracks in a Monte Carlo sample of light quark jets.

The $\log_{10}(p_T)$ distribution is used to define the probability $P(p_T)$ that the observed p_T of a given track is compatible with coming from a MB interaction:

$$P(p_T) = \frac{\int_{\log_{10}(p_T)}^{\inf} F(p_T) dp_T}{\int_{\log_{10}(0.5)}^{\inf} F(p_T) dp_T} \quad (1)$$

where $F(p_T)$ is the minimum bias track $\log_{10}(p_T)$ spectrum distribution obtained from the Monte Carlo simulation.

The probability that a vertex is consistent with a minimum bias interaction is given by

$$PMB = \prod \sum_{k=0}^{N-1} \frac{(-\ln \Pi)}{k!} \quad (2)$$

where Π is the product of the individual probabilities of the $N > 0$ tracks associated to the vertex. Only tracks with $p_T > 0.5$ are used for the calculation. We use the above definition instead of the simple product of track probabilities Π since it is independent of the number of tracks used in the calculation.

Figures 7 and 8 show the MB track probability distribution for minimum bias and hard scatter tracks in the simulation. The track probability for MB vertices is quite flat, as expected, indicating that the $F(p_T)$ distribution used correctly describes the MB characteristics. Any discrepancy between the $F(p_T)$ distribution used in the calculation and the actual distribution from the simulation would be seen as peaks at $P(p_T) = 0$ (if the discrepancy is at high p_T) or $P(p_T) = 1$ (if the differences are at low p_T).

The MB track probability for primary vertex interaction tracks peaks at 0, as expected, showing the incompatibility of primary vertex tracks with minimum bias tracks at high p_T . Figure 9 and 10 show the combined PMB vertex probability for minimum bias and hard scatter vertices in the simulation.

4 Performance in the Monte Carlo

We studied 4 different algorithms for primary vertex selection:

1. current p11 algorithm.
2. $\sum \log(p_t)$.
3. Vertex probability.
4. Interaction probability.

The last 3 algorithms are based on the z-vertex-clustering algorithm described in the section 2. The Interaction probability is calculated by using all tracks with transverse and longitudinal impact parameter with respect to the vertex within 0.15 cm and 0.5 cm respectively. Thus it includes tracks associated to “*split vertices*” and particles with long lifetime with high p_T .

The $\sum \log(p_t)$ algorithm is identical to the current p11 algorithm except for the z-vertex-clustering step that removes “*split vertices*”.

The performance in the Monte Carlo is analyzed in *Hddd* events with at least two reconstructed vertices separated by more than 2 cm of each other.

The primary vertex selection efficiency is defined as the number of correctly identified hard scatter vertices divided by the total number of Monte Carlo primary vertices. We require Monte Carlo vertices decaying to at least 3 charged particles within $|z| < 40\text{cm}$.

Next table shows the efficiencies for the 4 algorithms:

Algorithm	efficiency
p11	0.91
$\sum \log(p_t)$	0.95
Vertex probability	0.97
Interaction probability	0.99

Table 1: Primary vertex selection efficiency in Monte Carlo events.

There are 9 events for which the Interaction probability identifies a minimum bias interaction vertex as a primary vertex. Figure 11 shows the difference of probabilities between the selected vertex and the vertex with the closest probability. We can see that in these events, the 2 vertices look very similar each other in terms of track p_T distribution. The actual difference in the first bin of the histogram is of the order of 10^{-8} .

Figure 12 shows the interaction probability for these 9 events with 2 vertices with similar probabilities. In most cases the vertices are primary-like whereas in the remaining 3 events they seem to be compatible with minimum bias interactions.

5 Performance in the Data

In order to study the efficiency of the primary vertex selection in the data, we need to be able to identify hard scatter and min bias vertices in an unbiased way. For this purpose we selected di-muon events ($p_T > 2GeV/c$) where both muons were matched to global tracks and the z distance of closest approach between the 2 tracks was smaller than $500\mu m$. The average z position of the di-muon object (mostly J/Ψ s) was identified as the hard scatter vertex z position. We then study the primary vertex selection efficiency in events with at least 2 reconstructed vertices separated by more than 2 cm. Figure 13 shows that, in average, there are 1.2 vertices in this data sample.

Figure 14 shows the distribution of $\log_{10}(p_T)$ in the data and the simulation. Note that the track p_T turn-on is for $p_T > 0.5GeV$ or $\log_{10}(p_T) > -3$.

We could define the $F(p_T)$ distribution separately for data and Monte Carlo in order to account for the differences between simulated and real MB events, but since the agreement is reasonably good, we used the Monte Carlo $F(p_T)$ distribution also for the data.

Next table summarizes the primary vertex selection efficiency for the 4 different selection algorithms in the data and for $|z| < 40cm$.

Algorithm	efficiency
p11	0.88
$\sum \log(p_t)$	0.89
Vertex probability	0.92
Interaction probability	0.97

Table 2: Primary vertex selection efficiency in di-muon data events.

These results show that the vertex selection based on MB probabilities performs better than the $\sum \log(p_t)$ algorithm and the best performance is achieved when all tracks from the interaction (not only the tracks attached to the vertex) are used in the calculation. There are 2 main reasons for this improvement: on one hand, high p_T tracks from the hard interaction are always used, even if they don't get attached to the vertex due to resolution effects and on the other hand, high p_T particles with longer lifetime, are also used in the probability definition as these particles are most likely not attached to the primary vertex.

Figures 15, 16 and 17 show the distribution of number of attached tracks to the selected primary vertex for the different selection algorithms.

The Vertex probability algorithm significantly reduces the fraction of 2-track “*split vertices*” due to the fact that it selects the highest multiplicity vertex at each interaction z position. The remaining events at low track-multiplicity are minimum bias vertices misidentified as primary vertices. These misidentified events are removed when all tracks from the interaction are used (Interaction probability algorithm).

6 Conclusions

Vertex selection in p11 DØ reco version often selects several vertices coming from the same interaction consisting of 1 high multiplicity vertex and 1 or more “*split vertices*” made of 2 tracks with poor resolution. When the high p_T tracks of the events are attached to one of the “*split vertex*”, the current algorithm will select a “*split vertex*” as the primary vertex of the event.

We have a proposed a new algorithm for primary vertex selection that allow us to always choose the highest multiplicity vertex from disctints interactions using all the tracks (not only those attached to the vertex) to compute a MB/PV discriminant.

A new discriminant based on Minimum bias probability is presented and studied in both data and Monte Carlo. It has a better performance than the $\sum \log(p_t)$ discriminant currently being used in p11.

The MB vertex probability can also be used as part of other probabilistic algorithms such as the Missing Et significance. It will allow to re-vertex the jet and missing Et when 2 vertices have similar probabilities.

The next step is to re-process the electron plus jets and di-electron data samples with the new improved algorithm to increase the number of Z and W candidates in our data.

References

- [1] DØ Note 3906: “Primary Vertex Selection”, A. Schwartzman and M. Narain.

List of Figures

1	Track multiplicity of all selected vertices in the data.	10
2	Track multiplicity of the selected primary vertex in the data.	11
3	z distance between the selected primary vertex and all other vertices in the event.	12
4	Vertex χ^2 distribution of <i>split</i> and primary vertices.	13
5	Track multiplicity distribution of splitted and primary vertices.	14
6	Track p_T spectrum for Minimum bias and hard scatter vertices in the simulation.	15
7	Track probability distribution for minimum bias tracks.	16
8	Track probability distribution for primary vertex tracks.	17
9	Vertex probability distribution for minimum bias vertices.	18
10	Vertex probability distribution for primary vertices.	19
11	Probability difference between a misidentified primary vertex and the second vertex in the event. . .	20
12	Probability distribution of misidentified vertices. .	21
13	Vertex multiplicity distribution in di-muon data. .	22
14	Distribution of track $\log_{10}(p_T)$ for data (points) and the simulation (histogram)	23
15	Track multiplicity distribution for the current p11 selected primary vertex.	24
16	Track multiplicity distribution for the vertex probability algorithm.	25
17	Track multiplicity distribution for the interaction probability algorithm.	26

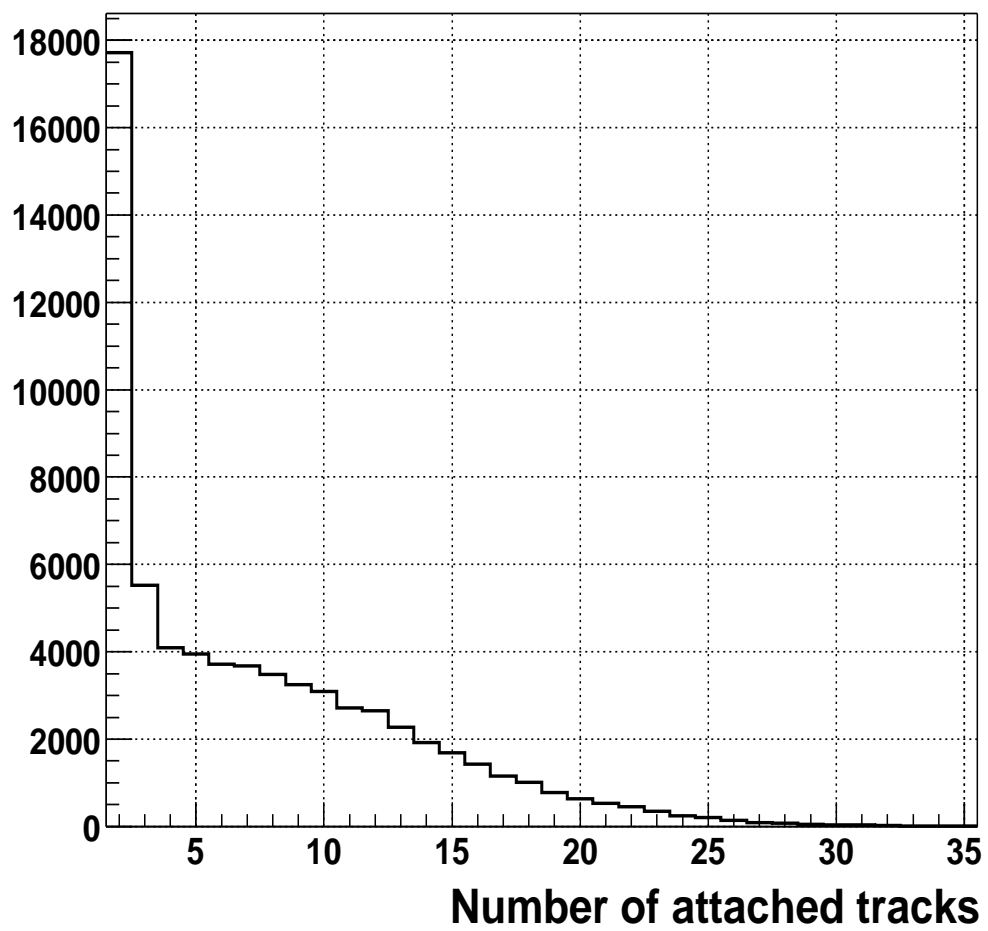


Figure 1: Track multiplicity of all selected vertices in the data.

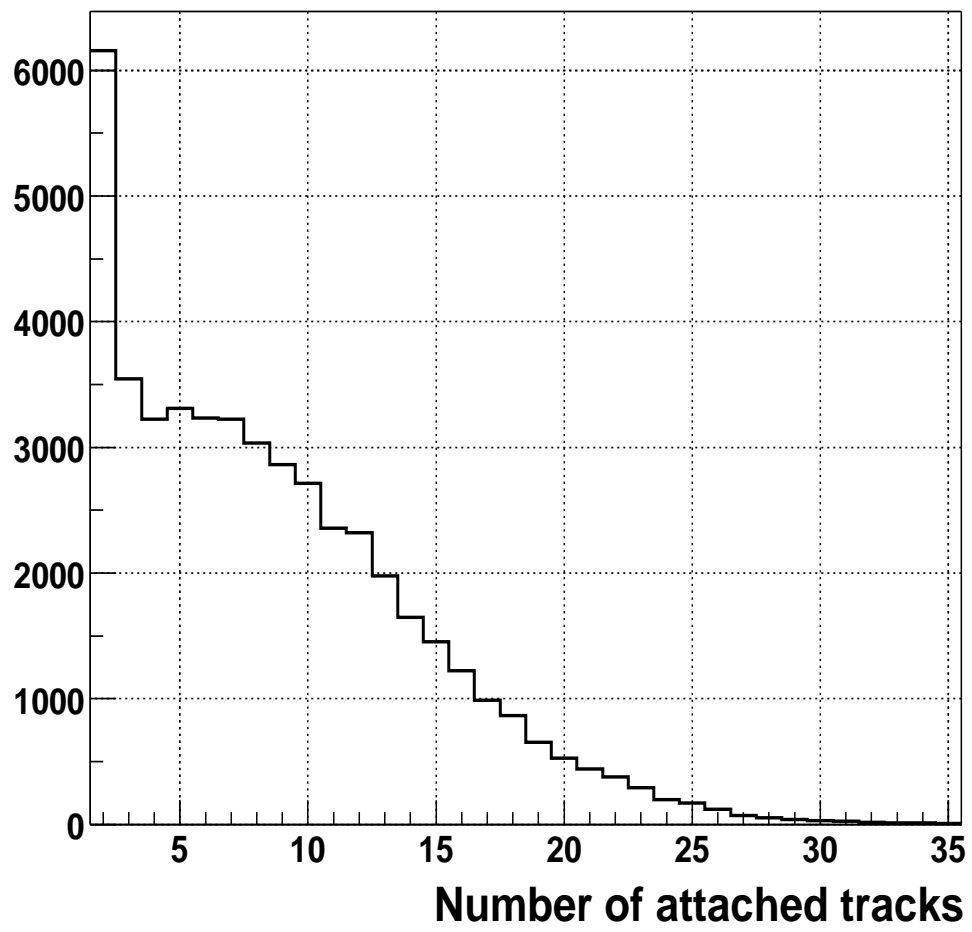


Figure 2: Track multiplicity of the selected primary vertex in the data.

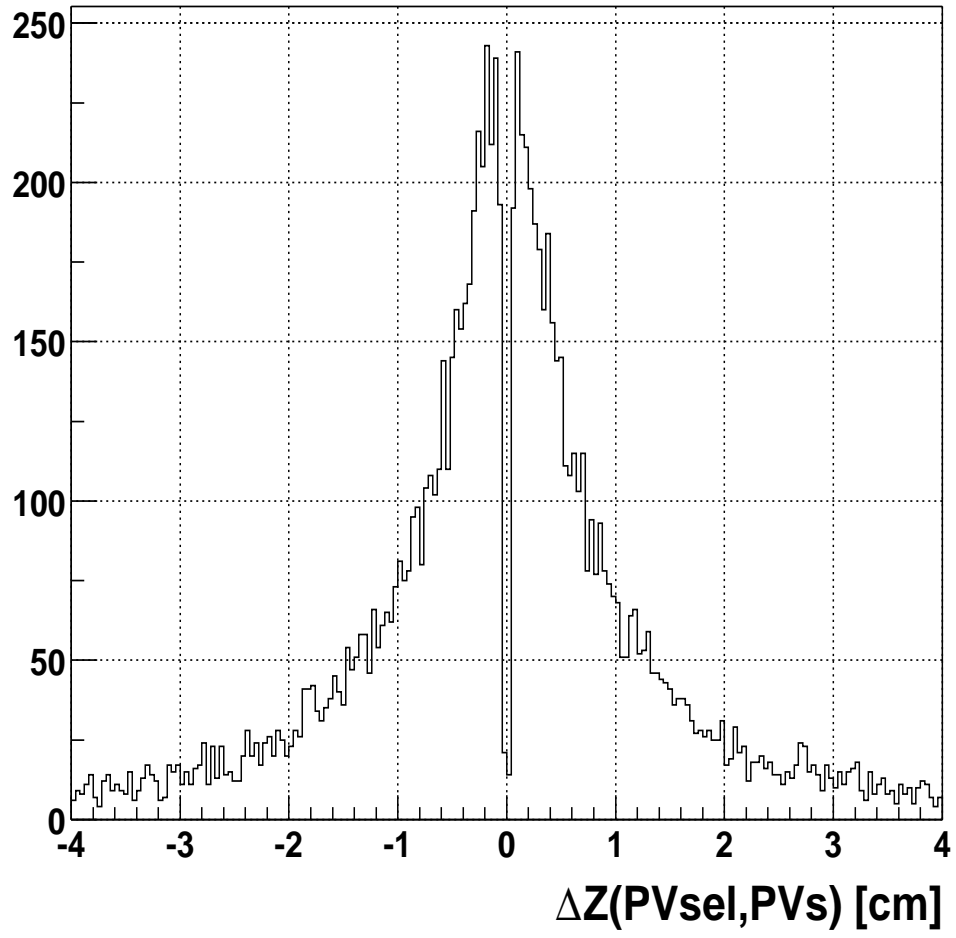


Figure 3: z distance between the selected primary vertex and all other vertices in the event.

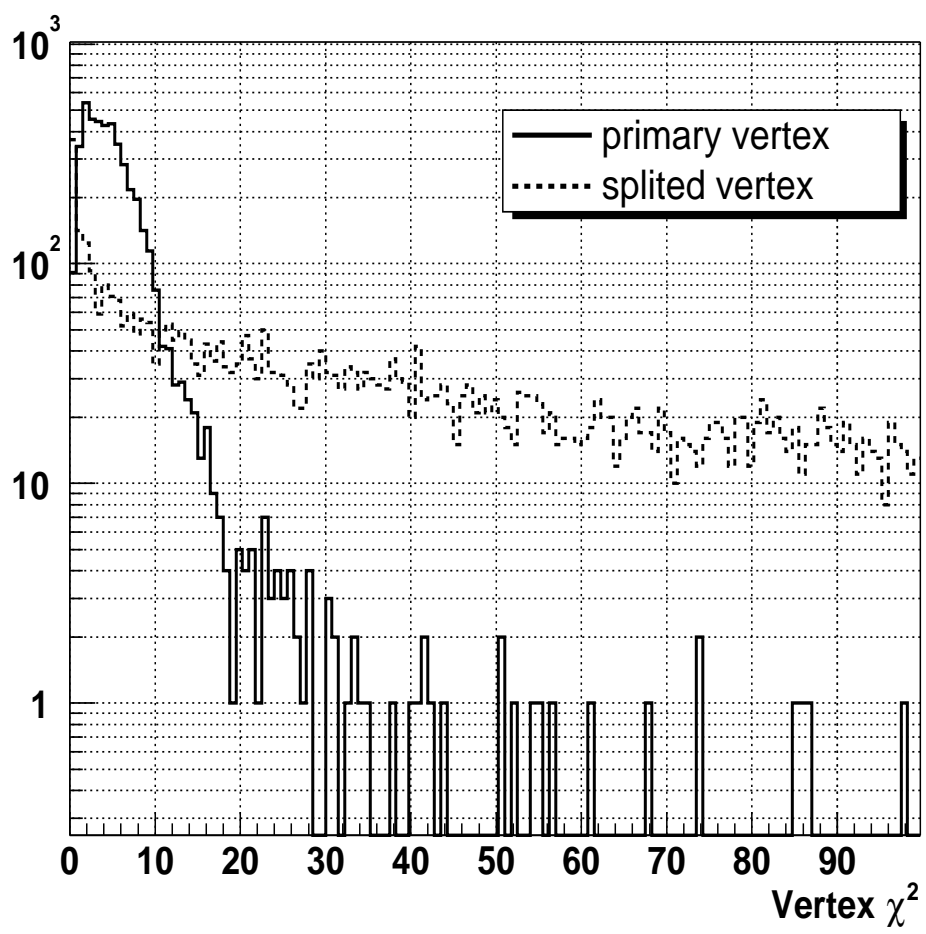


Figure 4: Vertex χ^2 distribution of *split* and primary vertices.

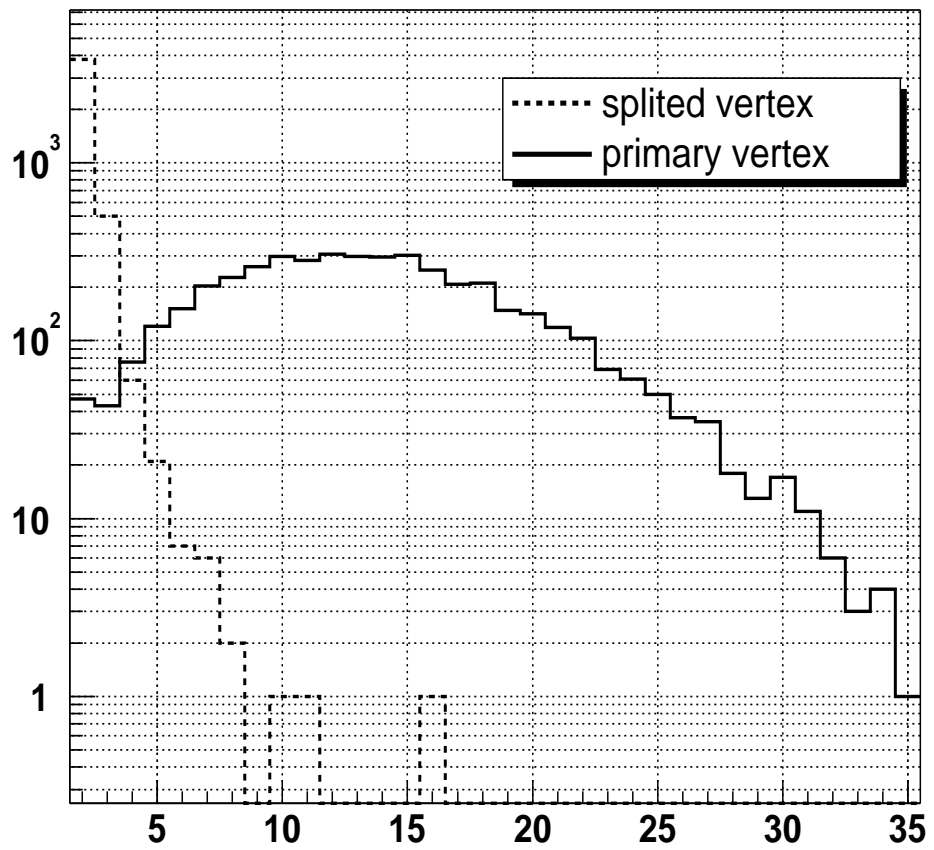


Figure 5: Track multiplicity distribution of splitted and primary vertices.

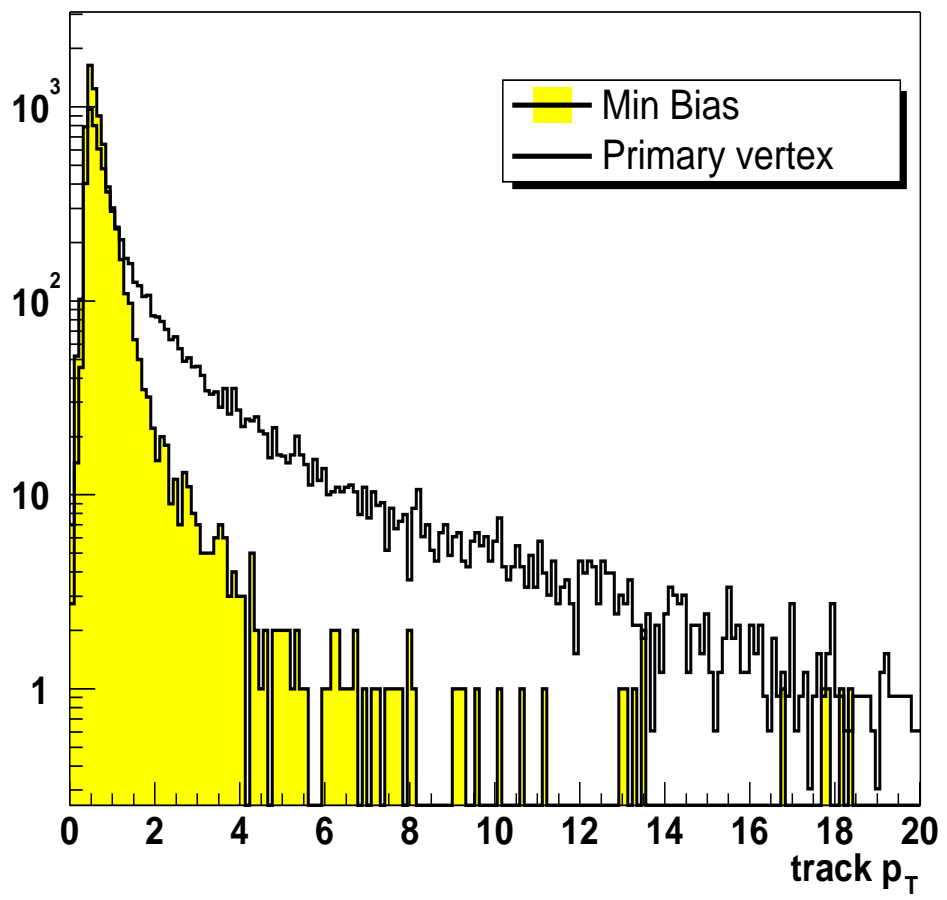


Figure 6: Track p_T spectrum for Minimum bias and hard scatter vertices in the simulation.

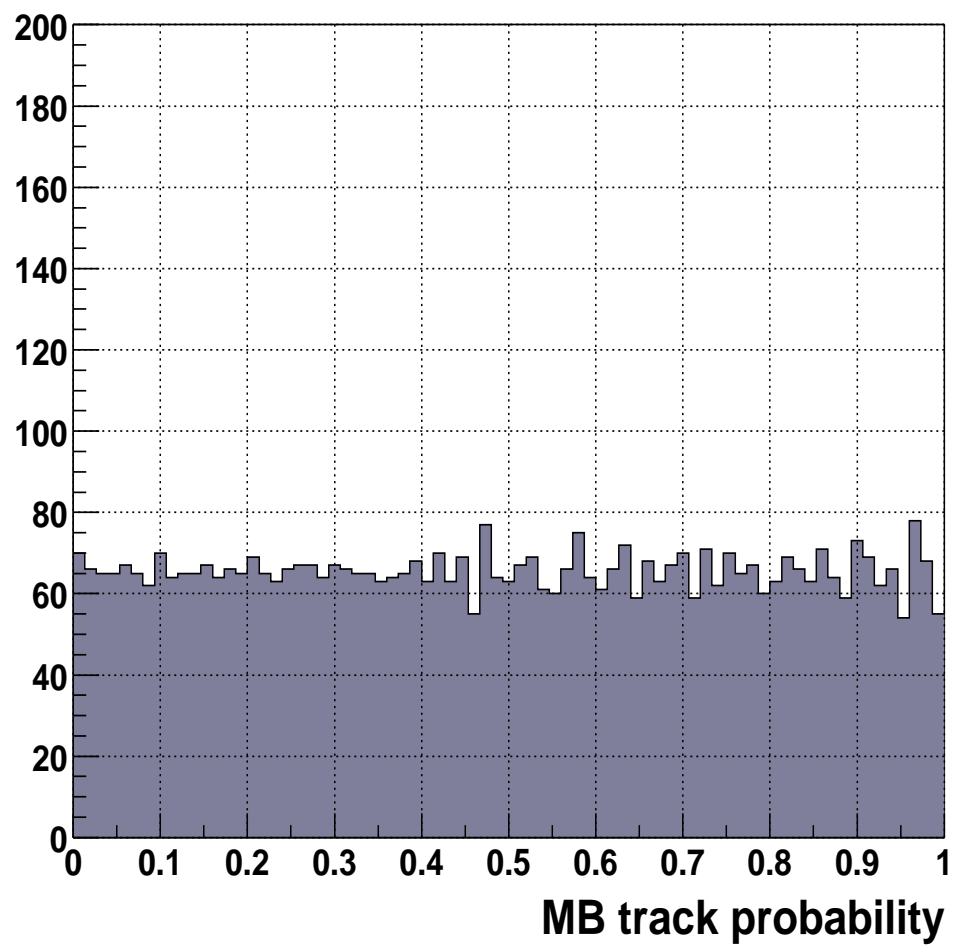


Figure 7: Track probability distribution for minimum bias tracks.

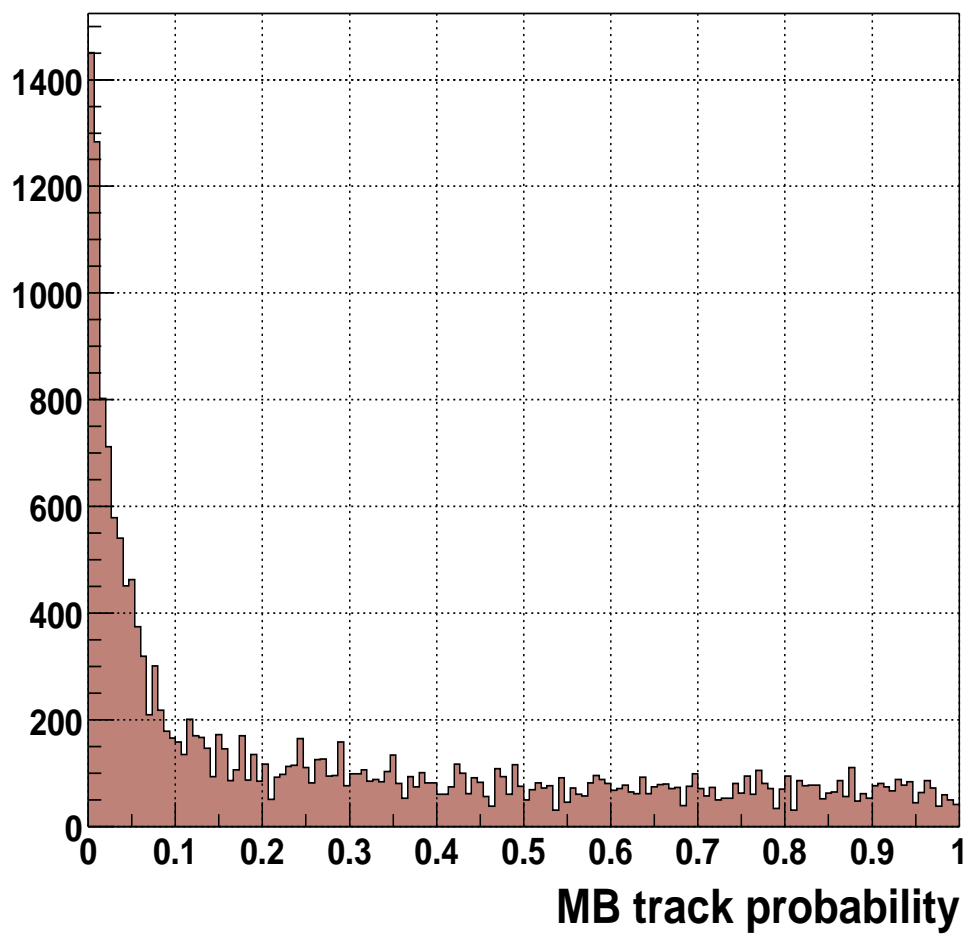


Figure 8: Track probability distribution for primary vertex tracks.

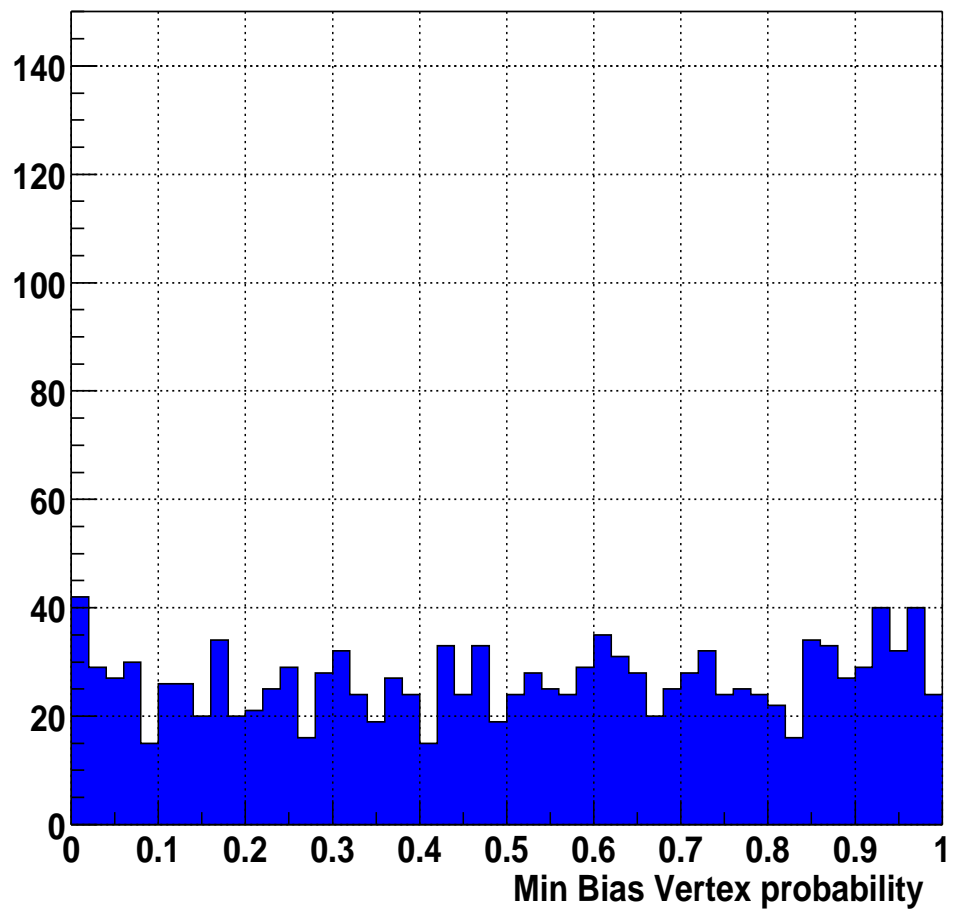


Figure 9: Vertex probability distribution for minimum bias vertices.

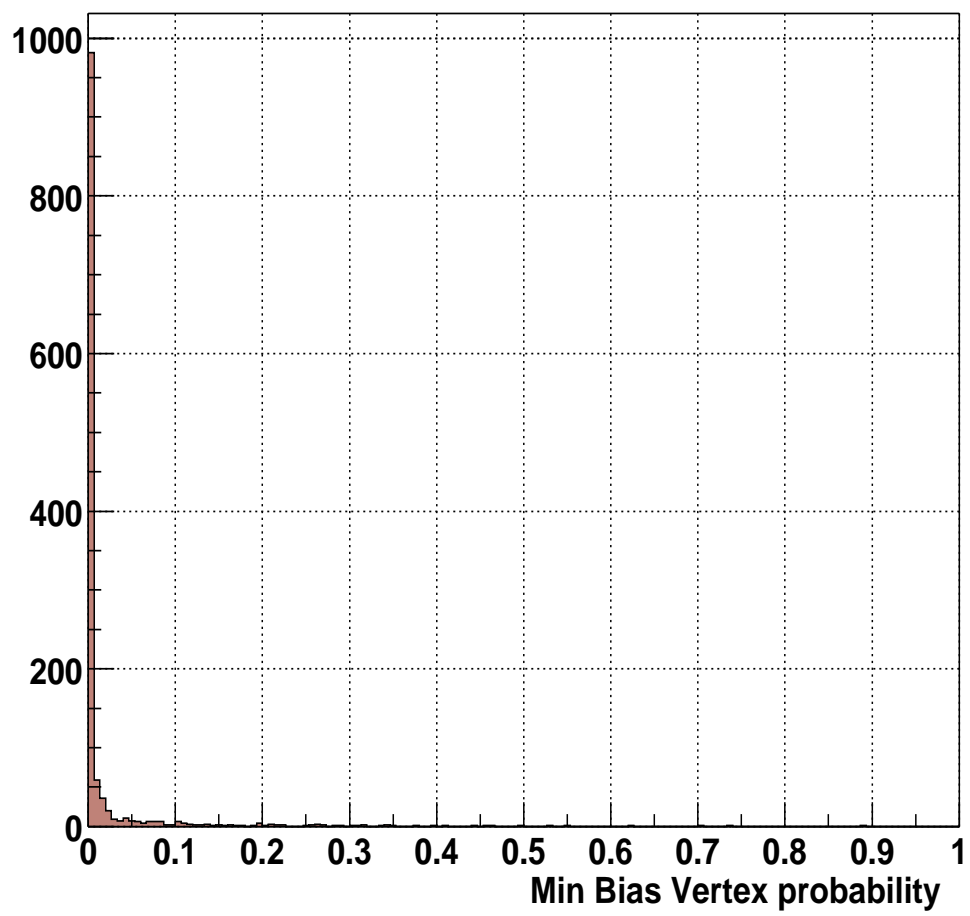


Figure 10: Vertex probability distribution for primary vertices.

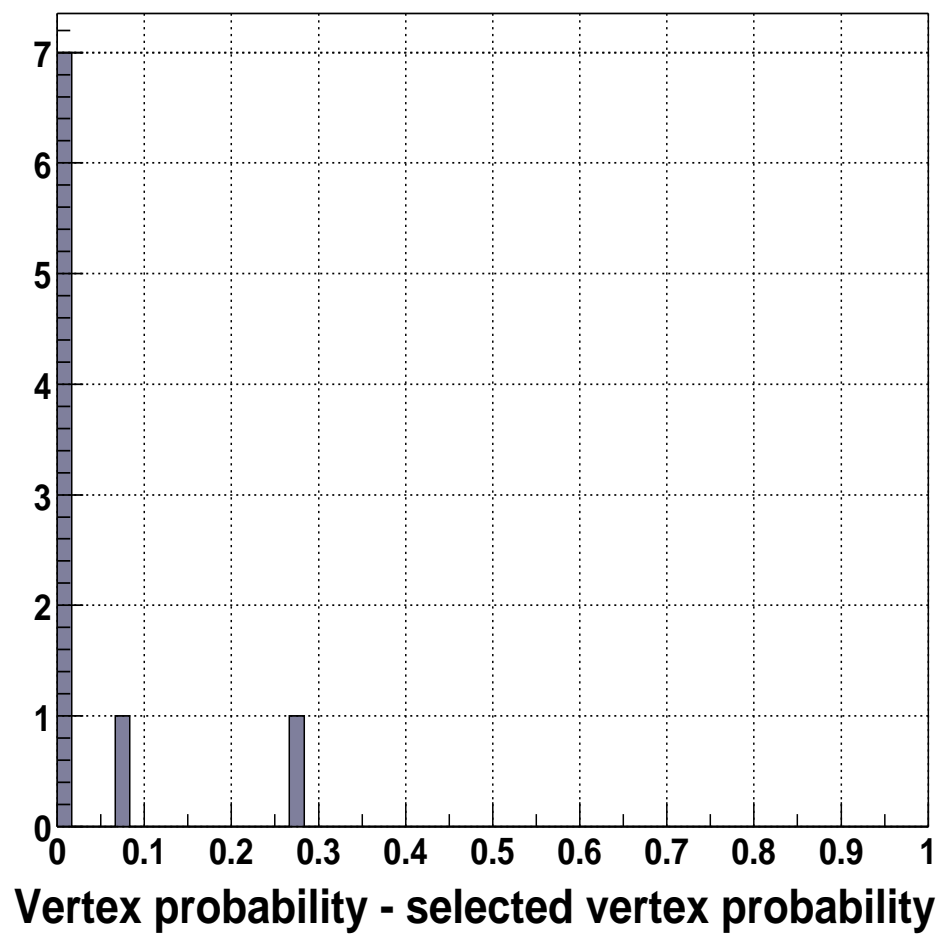


Figure 11: Probability difference between a misidentified primary vertex and the second vertex in the event.

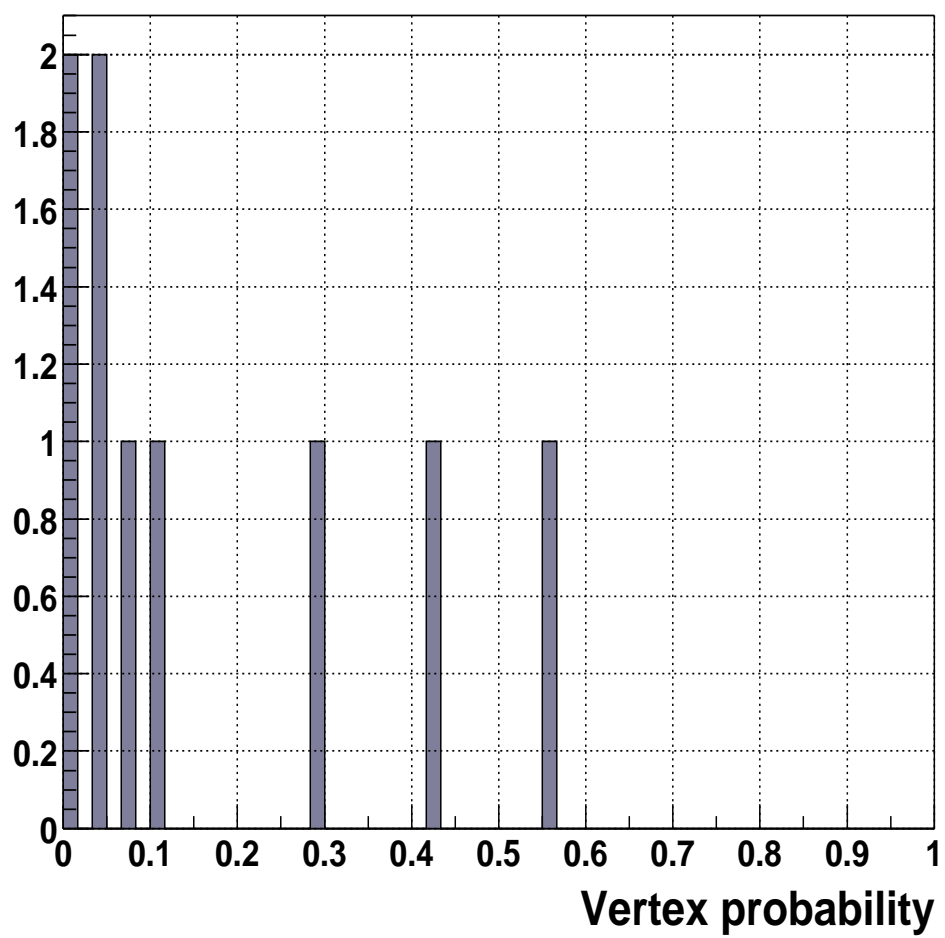


Figure 12: Probability distribution of misidentified vertices.

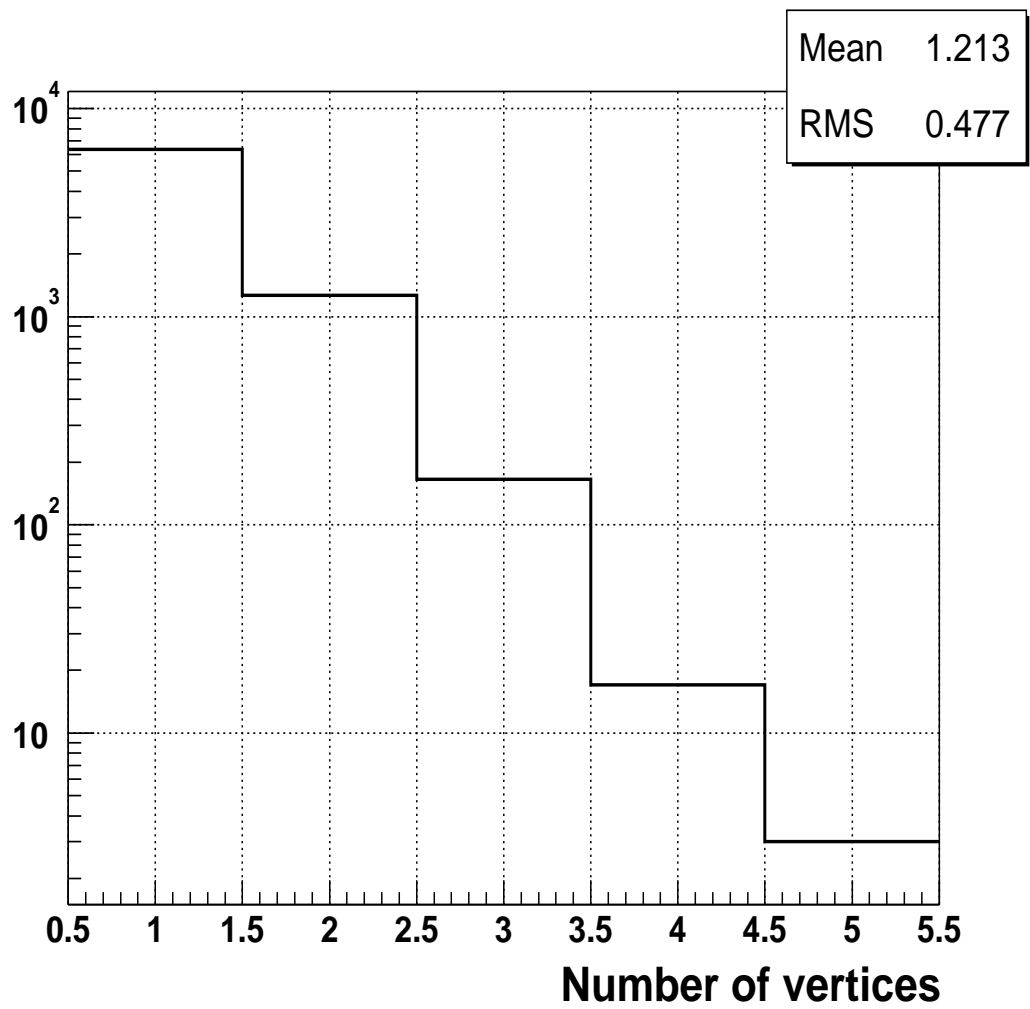


Figure 13: Vertex multiplicity distribution in di-muon data.

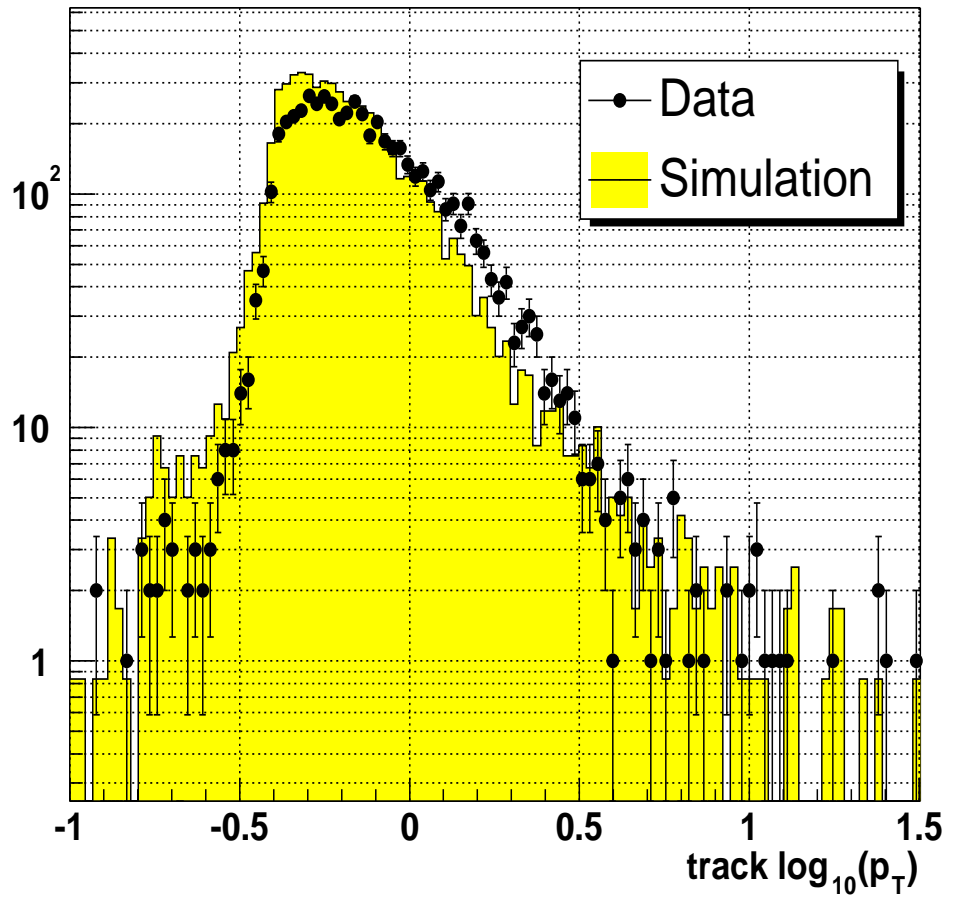


Figure 14: Distribution of track $\log_{10}(p_T)$ for data (points) and the simulation (histogram)

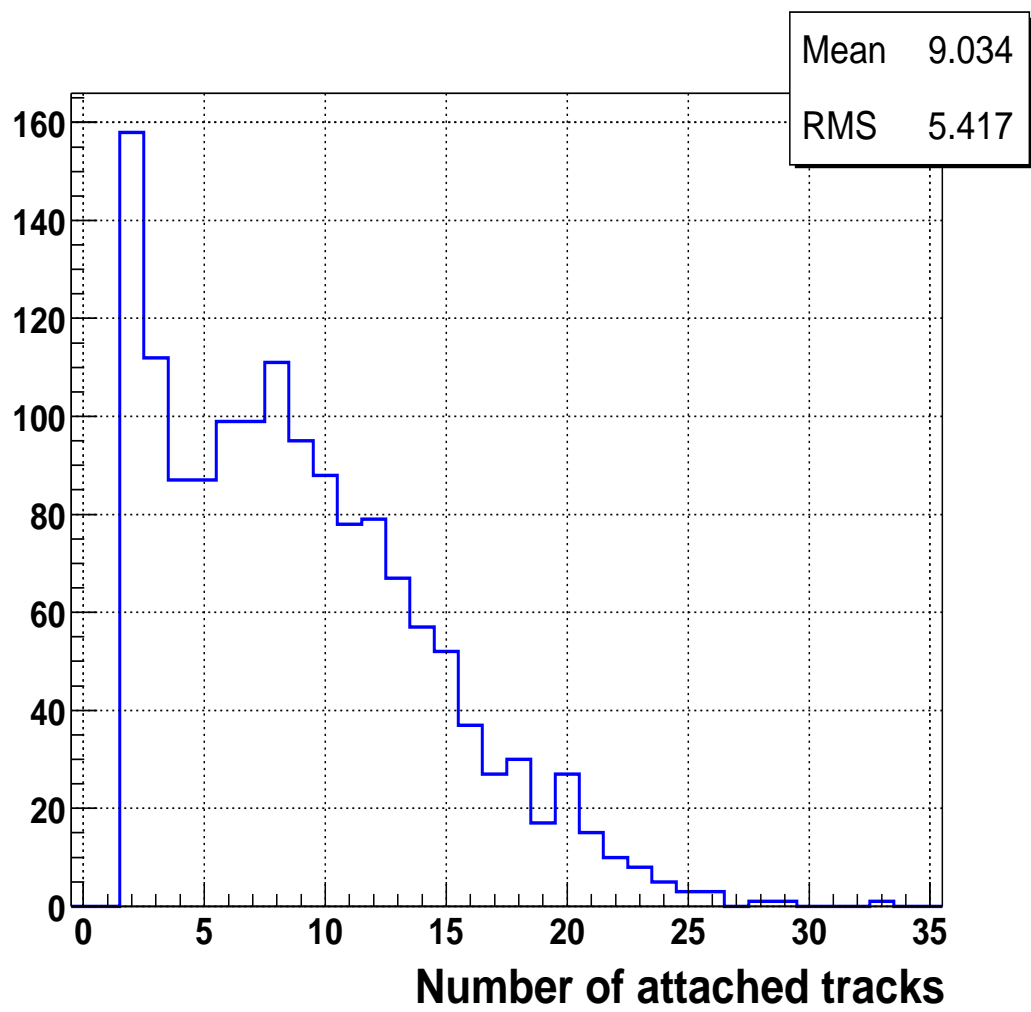


Figure 15: Track multiplicity distribution for the current p11 selected primary vertex.

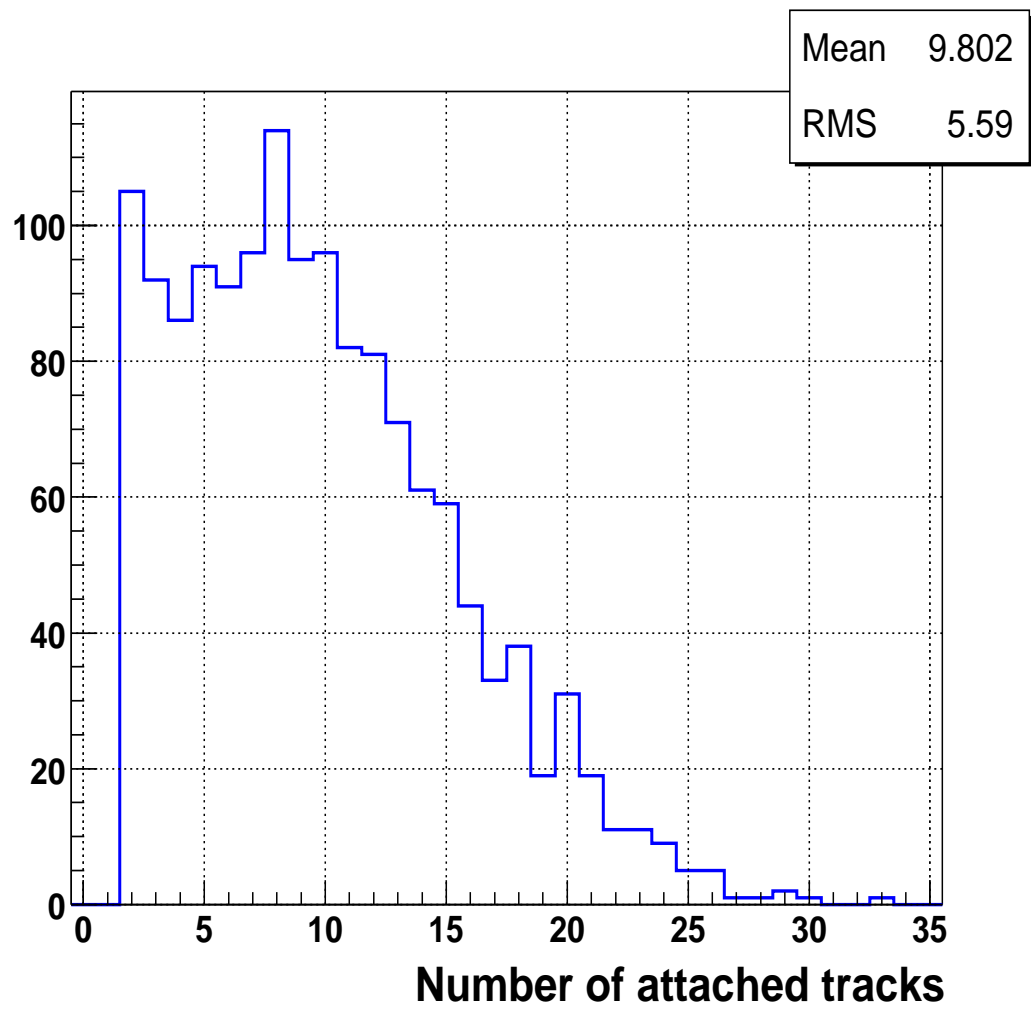


Figure 16: Track multiplicity distribution for the vertex probability algorithm.

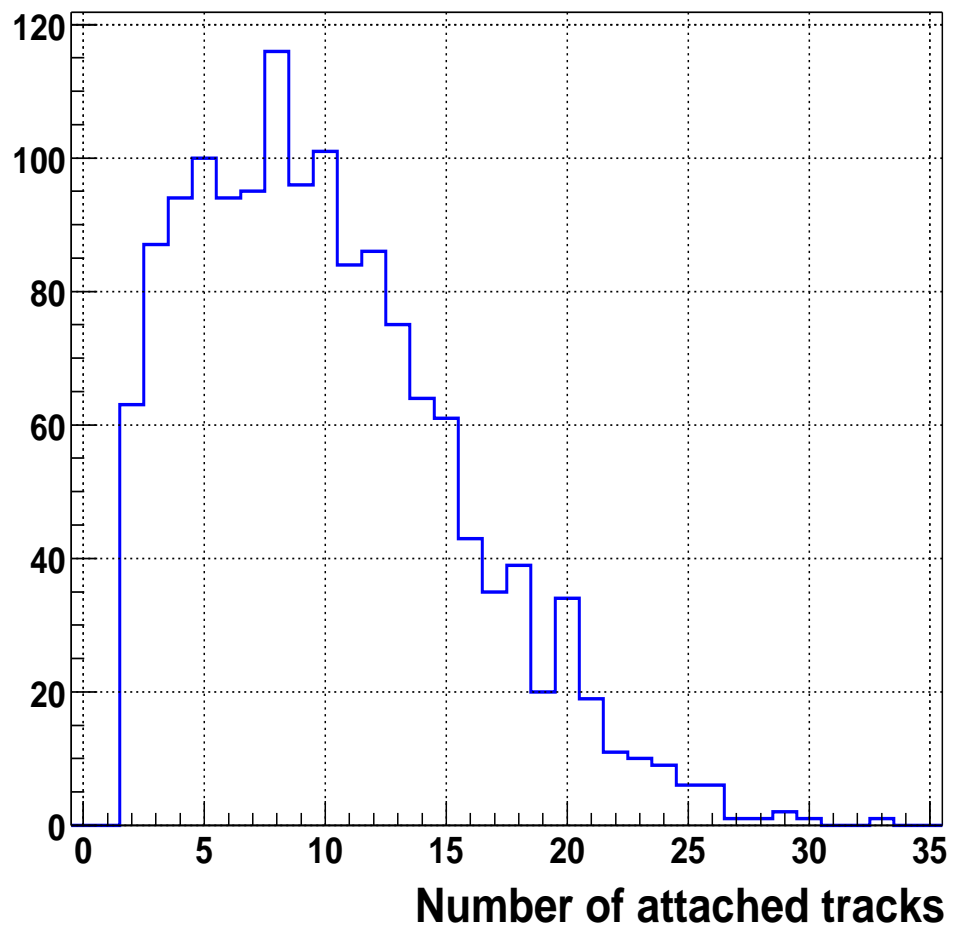


Figure 17: Track multiplicity distribution for the interaction probability algorithm.